Contagion at Work

Anna Houštecká
Dongya Koh
Raül Santaeulàlia-Llopis

December 2020

Barcelona GSE Working Paper Series

Working Paper nº 1225

Contagion at Work*

Anna Houštecká
Institute of Employment Research (IAB)

Dongya Koh
University of Arkansas

Raül Santaeulàlia-Llopis
UAB and BGSE

December 15, 2020

Abstract

Using nationally representative micro panel data on flu incidence from the Medical Expenditure Panel Survey in the United States, we show that employed individuals are on average 35.3% more likely to be infected with the virus. Wage earners are more likely to be infected than the unemployed by 30.1% and than individuals out of the labor force by 40.8%. Our results are robust to individual characteristics including vaccinations, health insurance and unobserved heterogeneity. Within the employed, we find an occupation-flu gradient—e.g. sales occupations show 34.1% higher probability of infection than farmers. As a potential mechanism behind this gradient, we study occupation-specific exposure to human contact interaction at work—a score that we construct based on O'NET occupational characteristics—which, as we show, determines flu incidence. All these effects increase with the aggregate flu incidence and are robust to firm size and across industries.

Keywords: Contagion, Flu, Employment, Unemployment, Occupations, Industry, Gradient, Exposure, Human Contact, Vaccines, Lockdown, Policy, Macroeconomics

JEL classification: E06, J01, J06

*We thank Chris Busch, Alex Ludwig and Ludo Visschers for their comments and suggestions at early stages of this project. We also thank the participants of the UAB Student and Alumni Webinar and of the Institute for Employment Research (IAB) Brownbag Seminar for their comments. Raül Santaeulàlia-Llopis acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Proyectos I+D+i 2019 Retos Investigacion PID2019-110684RB-I00 Grant and the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S).

1 Introduction

In pandemics in which a virus spreads via droplets, fomites and close *physical* human contact interaction (e.g., the 1918 influenza, SARS and Covid-19),[1] the public health response typically involves the enactment of economic lockdowns (e.g. Alvarez et al., 2020; Atkeson, 2020; Barro et al., 2020). The immediate goal of these policies is to reduce the spread of the virus generated from economic activity.[2] That is, at the core of lockdown policies is the trade-off between public health and economic activity. Therefore, the assessment of these policies inevitably requires knowledge about the contagion rate associated with economic activity. Unfortunately, since representative data on actual infections is rather limited, contagion rates at work (and elsewhere, for that matter) are not directly observable. At best, these contagion rates are estimated using proxies—i.e., potential infections—based on social contact and mixing patterns; see Mossong et al. (2008) and Klepac et al. (2020). In this manner, there is—as far as we know—no direct evidence on contagion rates constructed from actual incidence data for flu-like viruses.[3] Here, we exploit a rare opportunity that helps closing this gap by studying actual influenza (henceforth, flu) incidence from nationally representative micro panel data using the Medical Expenditure Panel Survey (MEPS). This survey collects individual flu data—together with a wide range of individual characteristics, including labor market variables—for a span of twenty years in the United States. We focus on studying contagion rates by employment status, occupations and industries. Several findings arise.

First, we find that employed individuals are more likely to be infected with the flu than the rest of the population controlling for year and season (month). Since the employed can differ from the non-employed by a set of individual characteristics, we estimate a benchmark employment gradient after controlling for age, gender, marital status, households size, and, importantly, medical conditions. In our benchmark specification, we find that the employed are 35.3% more likely to be infected with the virus than the non-employed. Unpacking the employed and non-employed, we find that wage earners are 8.9% more likely to be infected than the self-employed, 30.1% more likely to be infected than the unemployed, and 40.8% more likely to be infected than individuals out of the labor force. Additional controls on vaccination and access to health insurance do not alter our results.[4] Further, the panel structure of the MEPS allows for the

[1]For a description of the similarities and differences among these viruses, see this online discussion at the World Health Organization.

[2]For a dicussion on the effectiveness of these policies in containing the epidemics, see Aleman et al. (2020a).

[3]In the context of long-standing epidemics—in particular, for the human immunodeficiency virus (HIV) epidemic—there is substantially larger availability of nationally representative micro data; see Iorio and Santaeulàlia-Llopis (2016) and Aleman et al. (2020b).

[4]The CDC monitors influenza vaccine effectiveness (VE) every year through the US Flu VE Network. Across recent years the CDC reports VE of 60% in 2010-11, 47% in 2011-12, 49% in 2012-13, 52% in 2013-14, 19%

introduction of individual fixed effects in order to control for unobserved permanent differences in individual health. With individual fixed effects, the employment gradient increases to 38.4%.

Second, we find systematic differences in infection rates across occupations—and industries.[5] Conditional on our benchmark set of controls, we find that the occupation with the lowest flu incidence, "farming, fishing and forestry," shows a -7.3% lower incidence than our reference occupation group "production, transportation, and material moving operations." The occupation with the highest flu incidence, "sales and related occupations," has 24.4% higher flu incidence than the reference group. This implies an occupation-flu gradient of 34.1% between the occupation with the highest and lowest flu incidence. The occupation with the highest flu incidence is followed by "professional and related occupations", "management, business, and financial operations" and "office and administrative support" with, respectively, a 17.4%, 16.8% and 14.2% higher flu incidence than the reference group. Further, "service occupations" and "construction, extraction, and maintenance" show point estimates of flu incidence above the reference group though not significantly of 8% and 1%, respectively.

Next, we explore a potential mechanism for the occupation-flu gradient. Our idea is that occupations with more human contact interaction are also subject to larger contagion risk and, hence, flu incidence. With our data we can directly test whether this is the case (or not) and measure by how much. That is, we provide first direct evidence on how human contact interaction at work—a score that we construct based on occupation-specific work requirements—determines contagion. To do so, we use a set of O'NET descriptors and document how occupations differ in terms of the extent of human interaction required at work. In order to take into account all the specific forms of human interaction, we perform a principal component analysis through an eigenvalue decomposition of a covariance of all O'NET descriptors related to "Work Activities: Interacting with Others." Then, we merge our O'NET human contact score onto the MEPS occupation data so that we can explore the relationship between human contact interaction and flu incidence.[6] Our main result is that human contact interaction determines flu incidence.

in 2014-15, 48% in 2015-2016 and 40% in 2016-17. The medical literature suggests that lower VE is partially associated with a mismatch of the vaccine and the active strand of influenza. For example, in a recent meta-analysis of observational VE studies conducted in ambulatory care settings from 2004 to 2015, Belongia et al. (2016) show that VE against influenza B viruses was (on average) 54%, against A(H1N1)pdm09 viruses 61% and against H3N2 viruses 33%. Indeed, in the season with the lowest annual VE effectiveness over the past years (19%, in 2013-14) the CDC reports that the predominant virus was H3N2.

[5] MEPS provides individual data for eight occupation groups and thirteen industries. For the entire sample period, MEPS also specifies military specific occupations, which are by far the least affected by the flu. We also observe that 85% of those in military occupations had been vaccinated for the flu. We drop these military specific occupation and also "unclassifiable occupations", which together consist of less than 2% of our sample.

[6] To merge the scores of the O'NET occupations to the MEPS occupations we use several occupation codes crosswalks and the CPS occupational employment and individual weights. More details can be found in Appendix B.

Controlling for individual characteristics, we find that a one percent increase in the human contact score increases the probability of infection by 0.407 p.p. We also split the sample in years of high annual aggregate flu incidence (above median) and low annual aggregate flu incidence (below median). The human contact score significantly increases the probability of infection by a significant 0.532 p.p. in years of high annual aggregate incidence, whereas this effect is reduced in years of low annual aggregate incidence where the human contact score increases the probability of infection by a 0.282 p.p. Our results are robust to firm size—i.e. number of employees—and industry-fixed effects.

Finally, across industries, we find that the industry with the largest flu incidence is"education, health and social services" (EHSS). EHSS has 6 p.p more flu incidence than the industry with the lowest flu incidence, "natural resources." Potential reasons for the industry differentials in flu incidence include the within-industry occupation structure (e.g. there are more "professionals and related occupations" in EHSS than in the rest of the economy) as well as potential differences in the probability of infection by occupation across industries (e.g. "management, business, and financial operations" show higher probability of infection in EHSS than in the rest of the economy). With a simple decomposition exercise between EHSS and the rest of the economy, we show that differences in occupation structure explain 28.9% of the total difference in flu incidence across industries, whereas the remaining 71.1% is attributed to differences in the probability of being infected with the flu by occupation.

Related literature Our work relates to a strand of work in epidemiology that aims to estimate setting-specific contagion rates at work and other areas of human interaction. In this direction, Mossong et al. (2008) and Klepac et al. (2020) use primary data on human contacts to assess where infections potentially occur by constructing large-scale social mixing patterns for specific settings (e.g. home, work and school, among others). These studies, however, lack data on actual infections and might therefore not necessarily reflect how viruses actually spread within settings or across the population. We add to this literature by providing estimates on how a specific setting—work—contributes to actual flu incidence. Alternatively, some epidemiological studies, such as Ferguson (2005) which also focuses on the flu, are carried out under the assumption that the levels of contagion are approximately equal across settings (i.e., at work, school and the general community). This is also the case of Halloran et al. (2008) who, for example, write: *"In the absence of data to inform the choice, transmission in other contexts was arbitrarily partitioned to give levels of within-place transmission comparable with household transmission, namely 33% of transmission was assumed to occur in schools and workplaces, and 37% in the wider community (i.e., in contexts other than households, schools, and workplaces)."* In contrast, in our analysis we provide estimates of actual contagion at work.

Our work also relates to the macroeconomic literature that assesses lockdown policies in so far that policy assessment requires estimates of the contagion rate associated with economic activity. This type of policy analysis—extensively studied in the epidemiological literature on the flu (e.g. Ferguson, 2005; Halloran et al., 2008)—has recently regained interest due to the Covid-19 pandemic. In this direction, the current pandemic has generated a new and interesting growing literature on the aggregate and redistributive effects of lockdowns and their optimal implementation (see Alvarez et al. (2020), Atkeson (2020), Casares and Khan (2020), Eichenbaum et al. (2020), Farboodi et al. (2020), Fajgelbaum et al. (2020), Garibaldi et al. (2020), Glover et al. (2020) and Kaplan et al. (2020), among many others).[7] However, as far as we know, the current analyses of economic lockdowns to fight against the Covid-19 epidemic do not incorporate contagion rates at work (or other settings) informed by actual infections. In this context, our estimates on contagion rates at work by employment status, occupations and industries for the flu—in particular, for high aggregate incidence years—could be informative for the Covid-19 literature. Closest to our work is Almagro and Orane-Hutchinson (2020) that relates Covid-19 hospitalizations and occupation structure at the zip code level in New York. In contrast, in the context of the flu, we directly link occupations and incidence at the individual level without requiring any geographical (or other type of) aggregation. Finally, concurrent work also uses O'NET occupation information to provide alternative measures of occupation-specific human contact interaction at the occupation level (Mongey et al., 2020) or at the industry level (Azzimonti et al., 2020) in order to proxy for the exposure to the Covid-19 risk of infection. Again, our point of departure from these papers is that we directly link—at the individual level—our measure of occupation-specific human contact interaction with actual incidence. This allows us to assess, for the first time, the effects of human contact interaction at work on actual contagion.

2 Data

We use the Medical Expenditure Panel Survey (MEPS) and the Occupational Information Network (O'NET). We use MEPS to study individual flu incidence in relation with a wide range of labor market variables that include employment status, occupations, industries and firm size. We use O'NET to construct a measure of physical human contact by occupations in order to assess how the extent of physical human contact required by occupation relate to flu infections.

[7] This growing literature also includes testing and quarantine policies; see Berger et al. (2020), Obiols-Homs (2020) and Piguillem and Shi (2020), among others.

2.1 Medical Expenditure Panel Survey (MEPS)

The Medical Expenditure Panel Survey (MEPS) is a set of surveys of families and individuals, their medical providers and employers across the United States.[8] We use the the Household Component of the Medical Expenditure Panel Survey (MEPS-HC) which is a nationally representative survey of the U.S. civilian noninstitutionalized population. Data collection has been running since 1996 and the data set consists of yearly panels, in which data is collected for two calendar years for the same individual. Within those 2 years, MEPS conducts 5 rounds (interviews) on each individual.[9] The average number of families surveyed per year is 12,500 and the average number of individuals is 31,000. This implies a total amount of approximately 65,000 interviews per year and a total of approximately 4 million observations throughout. Further, note that the data from a particular round spans several months. To understand the seasonality of the flu in our data we create monthly data based on the interview month.

The Household Component includes data on health conditions, services and insurance together with information on individual employment (e.g., employment status, occupation, industry, hours worked, hourly wages, number of employed, whether or not an individual was self-employed, etc.) and other individual characteristics. The Medical Conditions files which are at the event-level give us information about all medical conditions that the individual reported in a particular round. This allows us to know in which round individuals report the flu and what other medical conditions they report in each round. In addition, the Preventive Care modules, which take place only in rounds 3 and 5, report information on the time since the individual got a flu shot.

Flu Incidence in MEPS On average, we find that the annual flu indicence, constructed as total number of individuals infected with the flu divided by the entire population in a given calendar year, is 1.6% between the years 1996 and 2016. The average flu incidence has large annual changes; see panel (a) of Figure 1. For example, in late 1990s the flu incidence reached average values for the entire population of approximately 2.4%. In contrast, in the late 2000s the average incidence was approximately 1%. The maximum average prevalence in our sample year reaches approximately 2.5% in 2012. The behavior of the flu incidence that we construct

[8] The data can be found on `https://www.meps.ahrq.gov/mepsweb/`.

[9] Each interview round collects information corresponding to a reference period, which is household-specific. The first round of each panel begins on the January 1 of the year where the panel started and ends at the date of the first interview. The reference periods for Rounds 2, 3, and 4 varied from household to household and covered the time between interview dates of the previous round and the current round. In our sample 80% of the first interviews were conducted between February and May of the first year, 80% if the second interviews were conducted between August and October of the first year. In the same way, 80% of the third and fourth interviews, conducted during the second year of the panel, were between January and April, and July and October, respectively. The last reference period ends on December 31 of the second year of the panel. A reference period thus corresponds, on average, to 5 months.

using MEPS is compared to that reported by the Center for Disease Control (CDC) in panel (b) of Figure 1. It is important to note that we use MEPS to document flu incidence, whereas the CDC's virologic surveillance for influenza reports the ratio between the number of positive tests divided by the total number of specimens tested—that is, a statistic conditional on testing. This explains the higher values in the CDC compared with MEPS. Nevertheless, the behavior of MEPS and CDC stastistics are similar after 2010.[10]

Although the flu is most common during the fall and winter, it is actually detected year-round.[11] The exact month at which the flu peaks in a year can change across years, but most of the flu activity peaks between October and March, see panel (c) in Figure 1.[12] In that figure, we show the patterns of flu incidence across months pooling the information from all our MEPS sample years, 1996-2016. The seasonal pattern that emerges is clear. The flu reaches its maximum incidence from October to March with a highest level of incidence at 2.56% in January and a lowest level of incidence close to 0.71% in June and July. Note that survey retrieves information with an average of five-month recall across interview rounds, which tends to mitigate the differences in incidence between months of high and low incidence.

The CDC estimates that during the 2018–2019 season there were 35.5 million people getting sick with influenza, 16.5 million people visiting a health care provider for their illness, 490,600 hospitalizations, and 34,200 deaths from influenza.[13,14] More than 46,000 hospitalizations occurred in children (aged $<$18 years); however, 57% of hospitalizations occurred in older adults aged $\geq$ 65 years. Older adults also accounted for 75% of influenza-associated deaths, highlighting that older adults are particularly vulnerable to severe outcomes resulting from an influenza virus infection. An estimated 8,100 deaths occurred among working age adults (aged 18–64 years), an age group that often has low influenza vaccination uptake. For this reason, we now also assess

[10]Unfortunately, we cannot reproduce the CDC statistic using MEPS data because we do not know who got tested for the flu in MEPS. In addition, while MEPS is constructed with a sample design that aims to be representative, the CDC data are not engineered to be necessarily representative. To start with, the sample tested is not representative. Further, the CDC data are collected using approximately 270 National Respiratory and Enteric Virus Surveillance System laboratories and 110 U.S. World Health Organization Collaborating Laboratories located in the U.S. These laboratories include public health laboratories and clinical laboratories and the CDC reports the test results from these laboratories separately only after 2015. Since public health laboratories often receive samples from a clinical laboratory to test the types of influenza virus, the results could be duplicated in prior years. Nevertheless, ignoring level differences, we find MEPS and CDC display similar trends after 2010 (see panel (b) in Figure 1), although there is substantial deviation between the two data sources before 2010.

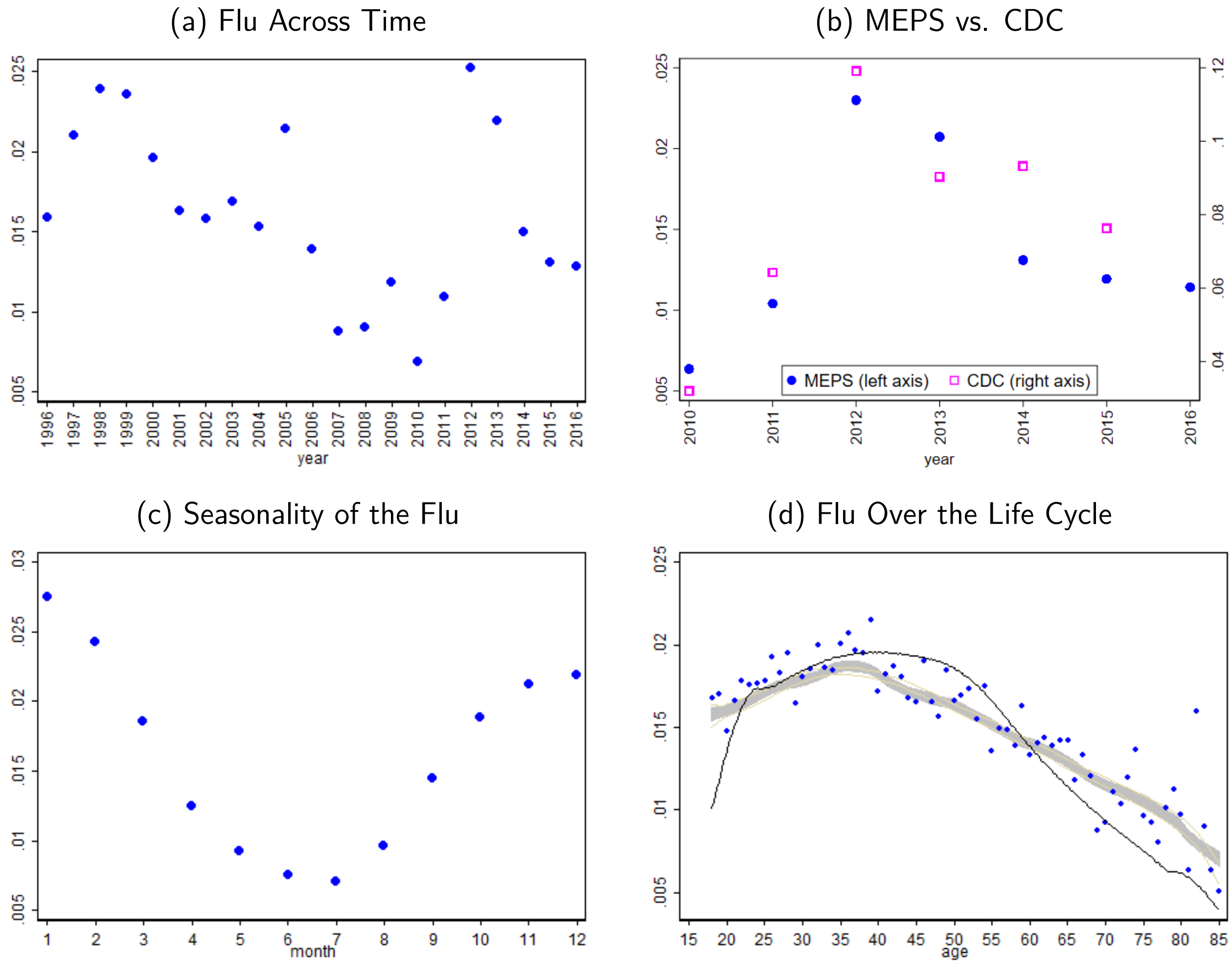
[11]This includes different strands of influenza of type A (e.g., H1N1pdm09, H3N2 and others) and of type B (e.g., Yamagata lineage, Victoria lineage, and others).

[12]Recently, the CDC reports similar peaks months of flu activity from a longer sample period, 1982-2018. See here: https://www.cdc.gov/flu/about/season/flu-season.htm.

[13]See https://www.cdc.gov/flu/about/burden/2018-2019.html.

[14]The number of influenza-associated illnesses that occurred last season was similar to the estimated number of influenza-associated illnesses during the 2012–2013 influenza season when an estimated 34 million people had symptomatic influenza illness.

Figure 1: Flu Incidence across Years, Seasons and Age, MEPS 1996-2016



Notes: This figure shows flu incidence, i.e. the total numbers of individuals infected with the flu divided by the total population, constructed using the MEPS 1996-2016 panels. In panel (a) we show how the flu incidence varies across years and in panel (b) we compare the flu incidence constructed from MEPS (left axis) with the number of specimens that tested positive for flu (conditional on testing) from the Center for Disease Control (CDC) clinical laboratories. In panels (c) and (d) we show, respectively, the (monthly) seasonality of flu incidence and the behavior of flu incidence over the life cycle. In panel (d) we include the empirical density of age in our MEPS sample (black line).

how flu incidence varies with age in our sample; see panel (d) of Figure 1 where we plot the age profile of the flu together with the age density in our sample. The flu incidence significantly differs by ages increasing from approximately 1.7% in the late teens to above 2% in the mid 30s to 40. After 40, the flu incidence relentlessly declines to values below 1.3% after retirement, and below 1% for individuals that are in the 70s. Despite the lower rate of flu infections in the older

Table 1: MEPS Sample Statistics

	Full Sample	Flu=1	Flu=0	Diff.	t
Employed	0.75	0.80	0.75	-0.05***	(-33.34)
Age	40.29	39.54	40.31	0.77***	(15.40)
Female	0.52	0.54	0.52	-0.02***	(-13.02)
Never Married	0.30	0.29	0.30	0.01**	(2.57)
HH Size	5.47	5.51	5.47	-0.03**	(-2.76)
N. Other Conditions	0.92	1.09	0.91	-0.18***	(-29.45)
Student Status	0.06	0.06	0.06	0.00	(0.57)
Flu Shot	0.49	0.49	0.49	-0.00	(-0.33)
N. Employees	101.23	107.09	101.13	-5.96***	(-9.13)
Observations	3,932,565	65,651	3,866,914		
Observations (Vaccines)	3,704,726				
% Of Total	(94.2%)	(1.67%)	(98.33%)		

Notes: This table shows the summary statistics of our benchmark sample (age 18-64), the sample of individuals infected with the flu (Flu=1) and not infected with the flu (Flu=0). We use sampling weights. The flu shot is observed for a subset of the individuals in the benchmark sample. The number of employees is set to zero for the non-working age population.

age groups, this is the population that is mostly affected with fatality cases. It is clear, however, that for individuals above 18 most of the infections occur to the working age population.

Benchmark Sample. To conduct our analysis, we use the first 20 panels of the MEPS, which span the years 1996-2016. In our benchmark sample we keep all individuals in the working age population, i.e., above 18 years old and below 65 years of age. This leaves us with 180,737 individuals and 3,932,565 monthly observations in total. Of these, 2,773,444 (70.5%)[15] observations were observed in employment, while the rest was not working.[16] In Table 1 we report the sample statistics for the full sample and across individuals who report a flu infection and those who report no flu infection. The average age is 40 years old, 52% are females and 30% have never been married, on average 5.7 people cohabitate and report less than one medical condition other than the flu (examples of other conditions include cancer, diabetes, sexually transmitted diseases or pneumonia).[17] Moreover, 6% are students, 50% of the observations had a flu shot, and the average number of employees is approximately 100.[18]

[15]Using sampling weights this corresponds to 75% of employed individuals in the sample

[16]In MEPS the employment status variables do not distinguish between unemployed workers and those out of the labor force. For individuals who have worked before, we observe the reason for not working which we can use to further decompose the non-working population.

[17]A detailed list of all medical conditions included in the MEPS can be found in Appendix 3 of https://meps.ahrq.gov/data_stats/download_data/pufs/h180/h180app3.html.

[18]We set the number of employees to zero for those who are not working.

Table 2: Work Activities: Interaction with Others

Name	
1. Assisting and Caring for Others	Providing personal assistance, medical attention, emotional support, or other personal care to others such as coworkers, customers, or patients.
2. Coaching and Developing Others	Identifying the developmental needs of others and coaching, mentoring, or otherwise helping others to improve their knowledge or skills.
3. Communicating with Persons Outside Organization	Communicating with people outside the organization, representing the organization to customers, the public, government, and other external sources. This information can be exchanged in person, in writing, or by telephone or e-mail.
4. Communicating with Supervisors, Peers, or Subordinates	Providing information to supervisors, co-workers, and subordinates by telephone, in written form, e-mail, or in person.
5. Coordinating the Work and Activities of Others	Getting members of a group to work together to accomplish tasks.
6. Developing and Building Teams	Encouraging and building mutual trust, respect, and cooperation among team members.
7. Establishing and Maintaining Interpersonal Relationships	Developing constructive and cooperative working relationships with others, and maintaining them over time.
8. Guiding, Directing, and Motivating Subordinates	Providing guidance and direction to subordinates, including setting performance standards and monitoring performance.
9. Interpreting the Meaning of Information for Others	Translating or explaining what information means and how it can be used.
10. Monitoring and Controlling Resources	Monitoring and controlling resources and overseeing the spending of money.
11. Performing Administrative Activities	Performing day-to-day administrative tasks such as maintaining information files and processing paperwork.
12. Performing for or Working Directly with the Public	Performing for people or dealing directly with the public. This includes serving customers in restaurants and stores, and receiving clients or guests.
13. Provide Consultation and Advice to Others	Providing guidance and expert advice to management or other groups on technical, systems-, or process-related topics.
14. Resolving Conflicts and Negotiating with Others	Handling complaints, settling disputes, and resolving grievances and conflicts, or otherwise negotiating with others.
15. Selling or Influencing Others	Convincing others to buy merchandise/goods or to otherwise change their minds or actions.
16. Staffing Organizational Units	Recruiting, interviewing, selecting, hiring, and promoting employees in an organization.
17. Training and Teaching Others	Identifying the educational needs of others, developing formal educational or training programs or classes, and teaching or instructing others.

Note: This table shows the descriptors included in the O'NET on Work Activities: Interacting with Others.

For the sample that reports a flu infection, we find that 78% of the individuals are employed, 54% are female, 29% were never married, 6% are students, and 50% had a flu shot in the past. As compared to the sample of individuals not infected with the flue, the proportion of employed individuals and females, the number of other conditions reported, the number of employees of the firm, and the average size of the household are significantly higher, whereas the average age and the proportion of never married individuals are significantly lower. The proportion of students and those who had a flu shot does not differ significantly between the flu and no-flu populations.

2.2 O'NET

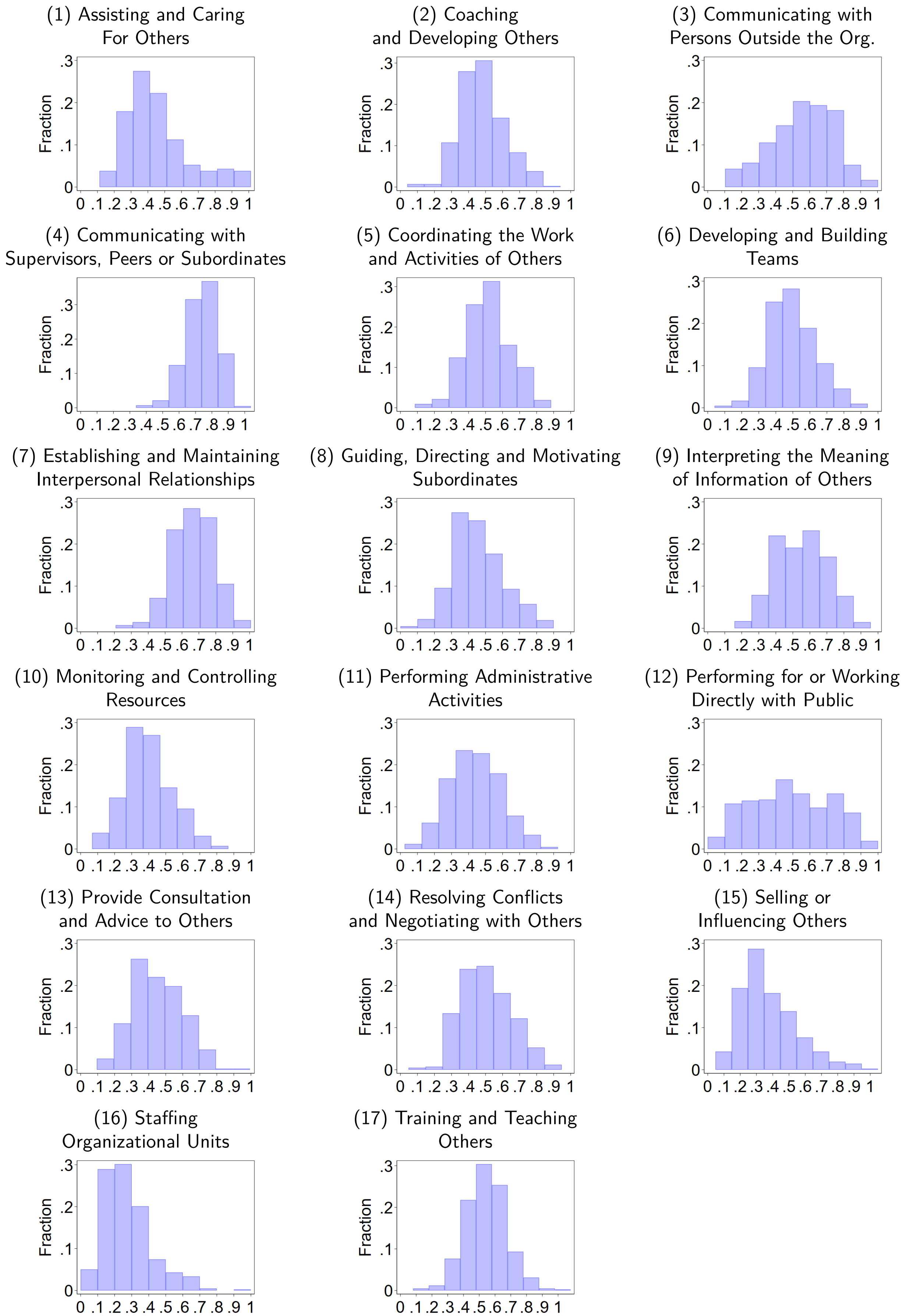
The Occupational Information Network (O'NET) is a database of worker and job characteristics in the United States. The database is constantly updated in terms of the number of occupations covered. The 24.2 Database used includes information for up to 774 occupations based on the Standard Occupation Classification of 2010 (SOC 2010). The information is gathered in a two-stage design; first, a random sample of businesses expected to employ workers in the occupations of interest is drawn and subsequently a random sample of workers in these occupations is drawn; then the individuals in this sample are surveyed by standardized questionnaires.

The O'NET questionnaires include questions about the importance of different abilities required at specific occupations. We are particularly interested in the O'NET database that regards with "Work Activities" and includes several indicators for the importance of "Interacting with Others", summarized in Table 2. The table shows a total of 17 different types of human interaction conducted at work. The answers to these questions are recorded from each respondent choosing a number between (and including) 1 and 5 (1 - Not important at all, 2 - Somewhat important, 3 - Important, 4 - Very important, 5 - Extremely important).

We show the distribution of each of the specific forms of human interaction in Figure 2. Each distribution is ordered with an importance scale that ranges from 1 (lowest) to 5 (highest). We normalize these categories to be between zero and one by subtracting one from the importance scale and dividing by four. We find substantial differences across forms of human interaction. The median normalized importance scale is lowest for work activities related to "Staffing Organizational Units" (0.24), "Selling or Influencing Others" (0.33), "Monitoring and Controlling Resources" (0.38) and "Assisting and Caring for Others" (0.41). The median normalized importance scale is highest for "Interpreting the Meaning of Information for Others" (0.55), "Establishing and Maintaining Interpersonal Relationships" (0.66), "Communicating with Persons Outside Organization" (0.70) and "Communicating with Supervisors, Peers, or Subordinates" (0.74).[19]

[19]These differences also arise for inequality statistics such as the standard deviation that is lowest for " Communicating with Supervisors, Peers, or Subordinates" (0.100), "Establishing and Maintaining Interpersonal Relationships" (0.126) "Training and Teaching Others" (0.129) and "Coaching and Developing Others" (0.131) and highest for "Selling or Influencing Others" (0.169) "Communicating with Persons Outside Organization" (0.181) "Assisting and Caring for Others" (0.191) "Performing for or Working Directly with Public" (0.232). These differences also arise in terms of the extent of skewness. The specific forms of human contact are typically positively skewed, except for three clearly negatively skewed work activities "Communicating with Supervisors, Peers, or Subordinates" (-0.654) "Communicating with Persons Outside Organization" (-0.333) and "Establishing and Maintaining Interpersonal Relationships" (-.223). The degree of positive skewness varies from relatively small coefficient values for "Performing for or Working Directly with Public" (0.036) to relatively high values for "Staffing Organizational Units" (1.118). We further conducted some Shapiro-Wilk normality tests. We can reject normality for all our specific forms of human interaction except for "Training and Teaching Others" and "Performing Administrative Activities." We note that while "Performing for or Working Directly with Public" had

Figure 2: Density (Fraction) of Specific Forms of Human Contact



Notes: This figure shows the distribution(fraction) of the importance of each of the specific forms of human interaction by occupation in the O'NET, based on the SOC 5-digit codes.

Finally, we note that the occupation and industry codes in the MEPS have changed in 2002. The MEPS provides a crosswalk between the pre-2002 and post-2002 classification only for the industry codes. Since after 2002, the occupational codes are condensed census codes[20] which we can map to the 25 census broad occupation groups and to the 2-digit SOC codes in the O'NET. We focus on the post-2002 period in our analysis of occupational differences in flu incidence.

3 Contagion at Work

We use the MEPS to study flu infection at work. First, we focus on how flu incidence varies with employment status in Section 3.1. Second, within the employed, we study how flu incidence varies across occupations in Section 3.2. In Section 3.3, we study how occupations differ by human contact interaction—a score that we construct based on occupation-specific work requirements and assess the effects of human contact interaction on flu incidence.

3.1 Employment Status and the Flu

We now study the employment gradient in flu incidence for individuals of 18 years of age and above.[21] Specifically, we run the following linear probability model specification:

$$flu_{it} = cons + \gamma e_{i,t} + \sum_t \omega_t \mathbf{1}_t + \sum_x \beta x_{it} + u_{it} \tag{1}$$

where flu_{it} is an indicator equal to one if individual i is diagnosed with the flu at month-year observation t and zero otherwise. If individuals are employed we set $e_{i,t}$ equal to one and zero otherwise and, therefore, γ captures the effects of employment on the probability of being infected with the flu. We capture the effects of years and seasons with a set of month-year dummies, $\mathbf{1}_t$. In our baseline specifications we also control for a set of individual characteristics that we denote with x_{it}. The set of individual characteristics include age (a cubic), gender, family size, marital status, and health status including a number of medical conditions reported in the same round as a proxy of the general health status of the individual. The residual is captured by u_{it}.

We first run our specification controlling for month-year effects but without individual controls. This gives us a measure of the unconditional effects of employment on the probability of being infected with the flu; see column (1) in Table 3. We find that, unconditionally, employed

a skewness coefficient close to zero, it shows a kurtosis (1.97) away from the normal distribution.

[20] The condensing rules can be found at https://meps.ahrq.gov/data_stats/download_data/pufs/ind_occ/occ3.shtml.

[21] In the MEPS, the labor market status tells us whether the individual has been employed during the round of reference or has not worked during the round. We do not consider the two other cases in which the individual reports that he 1) had a job to return to or 2) was employed sometime during the reference period.

individuals have higher flu infection rates than non-employed individuals. In particular, the estimated coefficient on employment implies a higher probability of infection of 0.461 percentage points (p.p.). This implies that employed and non-employed population we find that employed individuals are 33.3% more likely to get infected with the virus than the non-employed individuals.[22] The effect of employment on the probability of infection slightly increases after controlling individual characteristics to 0.483 p.p.; see column (2) in Table 1. This implies that, after controlling for individual characteristics, the employed are 35.3% more likely to get infected with the virus than the non-working population. Turning into the effects of individual characteristics, we find that flu incidence follows a hump shaped pattern over the life cycle captured by a cubic function of age; women have significantly higher chances of infection by 0.16 p.p.; higher number of medical conditions are associated with significantly higher flu incidence; and being a student also increases the odds of flu infection by 0.204 p.p. The size of the household does not significantly affect the probability of infection.

Vaccination and health insurance. Whether or not an individual gets infected is partially by flu vaccination—vaccine effectiveness against the flu is approximately 45% across recent seasons (2010-2017). It is important to control for vaccinations since the previous results could be a result of having better access to vaccination as employed individuals are more likely to have health insurance than non-employed individuals. Individuals in the MEPS were asked how long it had been since they had a flu shot.[23] This allows us to look at whether controlling for flu shots eliminates the employment gradient. We show in column (3) of Table 3 that this is not the case. In particular, the effect of employment remains large and significant. While the coefficient on ever getting vaccinated is not significant and positive,[24] the coefficient on the time since vaccine is positive and significant, suggesting that an additional year since the last time the individual had a vaccination is associated with 0.034 p.p. higher chances of having the flu; see column (4) of Table 3. In column (5) of Table 3 we control directly for whether an individual had health

[22]To see this, first we solve for the conditional flu prevalence of the non-employed individuals, x, in

$$\frac{\sum_i - \sum_i \mathbf{1}_{e_i}}{\sum_i} x + \frac{\sum_i \mathbf{1}_{e_i}}{\sum_i} (x + \gamma) \equiv 0.246x + 0.754 * (x + 0.461) = 1.73 \equiv \frac{\sum_i flu_i}{\sum_i}, \tag{2}$$

which implies that $x = 1.383$. Then, the employed are ((1.383+0.461)/1.383-1)*100=33.3% more likely to be infected than the non-employed population. Note that in equation (2) the term $(\sum_i - \sum_i \mathbf{1}_{e_i})/\sum_i$ is the share of non-employed in the population, $\sum_i \mathbf{1}_{e_i}/\sum_i$ is the share of employed in the population and $\sum_i flu_i/\sum_i$ is the aggregate (average) flu prevalence in the economy.

[23]In terms of vaccinations, MEPS provides information in rounds three and five. The question is retrospective and about how long ago individuals have been vaccinated last. This way, we can infer whether individuals were vaccinated (or not) in the previous rounds.

[24]This can be a result of the fact that a large number of individuals report that they had a flu shot during the last year. However, we do not observe when they had the shot and, therefore, this could be before or after getting infected with the flu.

insurance in the same round. The employment gradient is still present, while health insurance status is negatively associated to flu incidence, although not significantly.

Subjective Health Measures In our baseline specification, we control for the number of other medical conditions reported by the individual in the same round (apart from flu, if reported) to capture differences in individual health. We also run a specification where instead of using the number of other medical conditions reported we control for a self-reported general health variable; see column (6) in Table 3. The employment gradient in flu incidence increases slightly, and all other coefficients are close in magnitude to those in the baseline specification.

Metropolitan Area The MEPS includes data on whether or not an individual lives in a metropolitan area.[25] This is a potential confounder with employment, since employment can be larger in metropolitan areas, and at the same time the spread of influenza can be larger in metropolitan areas due to larger density of people off-work. However, the coefficient for employment remains significant even when we explicitly control for whether an individual resides in a metropolitan area; see column (7) in Table 3.

Wage-Earners, Self-employed, Unemployed and Out of the Labor Force In our analysis we have so far computed the gradient between the employed and the non-employed. Unpacking the non-employed we can group individuals between the unemployed and those out of the labor force.[26] In the same way, we also split the employed into self-employed and wage earners; Whe re-run our baseline specification using as reference group the individuals out of the labor force and with associated dummies for the unemployed, self-employed and the wage earners; see column (8) in Table 3. We find that the flu incidence of the unemployed is larger by 0.00104 p.p. but significantly different from that of the individuals that are out of the labor force. Compared with the individuals out of the labor force, the self-employed are 0.369 p.p. more likely to be infected with the flu, whereas the employed wage earners are 0.514 p.p. more likely to be infected with the flu. These results imply that wage earners are 8.9% more likely to be infected than the self-employed, 30.1% more likely to be infected than the unemployed, and 40.8% more likely to be infected than individuals out of the labor force.

Entire Population While we focus on the effects of employment within the working age population, we also assess the (average) employment gradient in the entire population. We now re-run our baseline specification using the entire population; see column (9) in Table 3. We also find an employment gradient where employed are 0.470 p.p. more likely to be infected with the flu.

[25] The MSA variable is only available up to 2015, thus, we run our regression on a smaller sample

[26] We define unemployed as those who report the reason for not working as 1) Could not find work, 2) On temporary layoff, and 3) Waiting to start new job.

Table 3: Employment Status and Flu Incidence

VARIABLES	(1) Year/Month FE	(2) Baseline	(3) Vaccines	(4) Vaccines	(5) Insurance	(6) Subj. Health	(7) MSA	(8) Multiple Categories	(9) Total Population
Employed	0.00461***	0.00483***	0.00478***	0.00491***	0.00470***	0.00519***	0.00476***	0.00514***	0.00470***
	(0.000417)	(0.000440)	(0.000452)	(0.000631)	(0.000452)	(0.000460)	(0.000488)	(0.000472)	(0.000404)
Self-Employed								0.00369***	
								(0.000770)	
Unemployed								0.00104	
								(0.00117)	
Flu Shot			0.000610						
			(0.000427)						
Time Since Vaccine				0.000356**					
				(0.000145)					
Health Insurance					-0.000188				
					(0.000494)				
Self-reported Bad Health						0.00341***			
						(0.000648)			
MSA							0.000657		
							(0.000559)		
Age		0.00135**	0.00135**	0.00187**	0.00182***	0.00140**	0.00131**	0.00138**	0.000589**
		(0.000587)	(0.000606)	(0.000853)	(0.000601)	(0.000587)	(0.000658)	(0.000592)	(0.000258)
Age^2		-3.22e-05**	-3.22e-05**	-4.51e-05**	-4.43e-05***	-3.38e-05**	-3.08e-05*	-3.24e-05**	-1.26e-05**
		(1.45e-05)	(1.50e-05)	(2.08e-05)	(1.50e-05)	(1.46e-05)	(1.63e-05)	(1.47e-05)	(5.14e-06)
Age^3		2.24e-07*	2.22e-07*	3.22e-07**	3.22e-07***	2.42e-07**	2.08e-07	2.23e-07*	6.57e-08**
		(1.15e-07)	(1.19e-07)	(1.63e-07)	(1.19e-07)	(1.15e-07)	(1.28e-07)	(1.16e-07)	(3.20e-08)
Female		0.00167***	0.00156***	0.00154***	0.00140***	0.00220***	0.00169***	0.00173***	0.00180***
		(0.000407)	(0.000419)	(0.000589)	(0.000422)	(0.000406)	(0.000450)	(0.000411)	(0.000354)
Never Married		-0.000573	-0.000764	-0.000321	-0.000708	-0.000686	-8.40e-05	-0.000513	-0.000561
		(0.000569)	(0.000583)	(0.000811)	(0.000581)	(0.000570)	(0.000642)	(0.000575)	(0.000543)
HH Size		-4.71e-05	-9.48e-05	-7.75e-05	-8.49e-05	-7.06e-05	-3.89e-05	-3.27e-05	-2.48e-05
		(7.03e-05)	(7.19e-05)	(0.000104)	(7.27e-05)	(7.01e-05)	(7.86e-05)	(7.09e-05)	(6.29e-05)
N. Other Conditions		0.00150***	0.00145***	0.00156***	0.00165***		0.00121***	0.00147***	0.00128***
		(0.000159)	(0.000164)	(0.000208)	(0.000169)		(0.000176)	(0.000161)	(0.000125)
Student Status		0.00204*	0.00215*	0.00102	0.00266**	0.00223*	0.00199	0.00212*	0.00143
		(0.00119)	(0.00123)	(0.00180)	(0.00119)	(0.00119)	(0.00135)	(0.00120)	(0.00111)
Constant	0.0257***	0.00646	0.00652	-0.00124	0.0230***	0.00735	0.00609	0.00561	0.0149***
	(0.000895)	(0.00762)	(0.00788)	(0.0113)	(0.00827)	(0.00762)	(0.00857)	(0.00771)	(0.00408)
Observations	3,932,565	3,932,565	3,704,726	1,746,050	3,343,721	3,932,565	3,212,569	3,880,624	4,737,274
R-squared	0.005	0.006	0.006	0.006	0.005	0.005	0.006	0.006	0.006
Year/Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R2	0.00513	0.00550	0.00554	0.00550	0.00527	0.00535	0.00624	0.00552	0.00558

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 1996-2016. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Employment Status and Flu Incidence: Individual Fixed Effects

VARIABLES	(1) Baseline	(2) Indiv. FE	(3) Indiv. FE + Vaccine
Employed	0.00483***	0.00496***	0.00515***
	(0.000440)	(0.00110)	(0.00117)
Flu Shot			-0.00145
			(0.00101)
Age	0.00135**	0.00297	0.00284
	(0.000587)	(0.00302)	(0.00314)
Age^2	-3.22e-05**	-0.000117	-0.000112
	(1.45e-05)	(7.70e-05)	(8.05e-05)
Age^3	2.24e-07*	1.03e-06*	9.85e-07
	(1.15e-07)	(6.17e-07)	(6.47e-07)
Female	0.00167***		
	(0.000407)		
Never Married	-0.000573	-0.000649	-0.000236
	(0.000569)	(0.00253)	(0.00270)
HH Size	-4.71e-05		
	(7.03e-05)		
N. Other Conditions	0.00150***	-0.00153***	-0.00139***
	(0.000159)	(0.000205)	(0.000210)
Student Status	0.00204*	0.00317*	0.00423**
	(0.00119)	(0.00189)	(0.00186)
Constant	0.00646	0.0707*	0.0595
	(0.00762)	(0.0396)	(0.0408)
Observations	3,932,565	3,932,565	3,704,726
R-squared	0.006	0.005	0.006
Year/Month FE	Yes	Yes	Yes
Individual FE	No	Yes	Yes
Pseudo R2	0.00550	0.00548	0.00551
Number of id		180,737	177,586

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 1996-2016. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Individual Fixed Effects. Given the panel structure of the MEPS and the fact that we have enough variation in employment status within individuals we also run a specification including individual fixed effects in order to control for potential unobserved individual characteristics. The coefficient stays remarkably stable even if we use within-individual variation to study the relationship between employment status and flu incidence. If anything, controlling for unobserved heterogeneity increases the employment gradient to 0.496 p.p.; see column (2) of Table 4. If in addition to controlling for unobserved heterogeneity we also control for vaccinations we find even a slightly larger employment gradient of 0.515 p.p.; see column (3) of Table 4. This implies that controlling for unobserved heterogeneity and vaccinations, the employed have a 38.4% higher probability of being infected with the flu than the non-employed individuals. It is interesting to note that the sign of the coefficient of the number of other conditions changes. This points to the

Table 5: Employment Status and Flu Incidence: High versus Low Aggregate Incidence

VARIABLES	(1) All years	(2) High Incidence	(3) Low Incidence
Employed	0.00483***	0.00574***	0.00393***
	(0.000440)	(0.000701)	(0.000498)
Age	0.00135**	0.00110	0.00153**
	(0.000587)	(0.000906)	(0.000681)
Age^2	-3.22e-05**	-2.24e-05	-4.01e-05**
	(1.45e-05)	(2.25e-05)	(1.69e-05)
Age^3	2.24e-07*	1.23e-07	3.09e-07**
	(1.15e-07)	(1.78e-07)	(1.34e-07)
Female	0.00167***	0.00239***	0.000963**
	(0.000407)	(0.000629)	(0.000480)
Never Married	-0.000573	0.000469	-0.00159**
	(0.000569)	(0.000887)	(0.000666)
HH Size	-4.71e-05	5.17e-05	-0.000128
	(7.03e-05)	(0.000114)	(8.11e-05)
N. Other Conditions	0.00150***	0.00112***	0.00191***
	(0.000159)	(0.000243)	(0.000198)
Student Status	0.00204*	0.00139	0.00256*
	(0.00119)	(0.00184)	(0.00135)
Constant	0.00646	0.0101	0.00465
	(0.00762)	(0.0117)	(0.00878)
Observations	3,932,565	1,995,272	1,937,293
R-squared	0.006	0.006	0.003
Year/Month FE	Yes	Yes	Yes
Pseudo R2	0.00550	0.00553	0.00329

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 1996-2016. High incidence years are those where flu incidence was higher than 1.67%. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1

fact that variation in the number of other conditions across individuals summarizes well differences in permanent health across individuals, and those more prone to having other medical conditions also catch the flu. However, when using time variation within individual, flu rounds are associated with a lower number of other reported conditions.[27] Including vaccinations, while insignificant, we find a negative relationship between vaccinations and flu incidence within individuals.

Aggregate Flu Incidence. Across years, the median annual aggregate flu incidence was of 1.67%. Using the median aggregate flu incidence as a threshold, we split our sample between years above and below the median aggregate incidence which we define, respectively, as high

[27]Since for each medical condition in the MEPS we observe the first round it was recorded, it is possible that an individual reports less new conditions in the same round as the flu as they are less likely to contract other contagious illnesses and less likely to get preventive care for other conditions and get diagnoses of longer term conditions such as cancer or diabetes.

and low incidence years. We find that the employment gradient of the flu is increasing with the aggregate flu incidence in the economy. In high aggregate flu incidence years the employed have higher infection rates than the non-employed by 0.574 p.p.; see column (2) in Table 5. This estimate is of 0.393 p.p. in years with low aggregate flu incidence; see column (3) in Table 5. Further, the difference in the employment gradient of the flu is significant between high and low aggregate incidence years.

3.2 Flu Incidence by Occupations

In this section we report the flu incidence by occupation. The MEPS categorization of occupations changed in 2002 and we thus restrict our attention to the 2002-2016 sample. After 2002, the occupational groups are condensed census codes and correspond to the following eight occupations: management, business, and financial operations; professional and related occupations; service occupations; sales and related occupations; office and administrative support; farming, fishing and forestry; construction, extraction, and maintenance; and production, transportation, and material moving operations.[28] The two largest occupations are "professional and related occupations" and "service occupations" that, respectively, account for 23% and 16.4% of the total employment in the MEPS for 2002-2016. In terms of employment size, they are followed by "management, business, and financial operations","office and administrative support" and "production, transportation, and material moving operations" that account for, respectively, 15.6%, 13.2% and 12.5% of total employment. The occupations of "sales and related occupations" and "construction, extraction, and maintenance" represent, respectively, 9.7% and 8.9% of total employment. The occupation with the lowest share of employment is "farming, fishing and forestry" that represents 0.7% of total employment.

We now estimate how flu incidence differs by occupations. Precisely, we compute the occupation-specific effects on flu incidence estimates, β_{occ}, from following regression specification,

$$flu_i = cons + \sum_{occ \neq 1} \beta_{occ} \mathbf{1}_{occ} + \sum_{t} \omega_t \mathbf{1}_t + u_{it}, \tag{3}$$

where we are using "production, transportation and material moving operations" as reference occupation indexed as 1, and we control for year and month. Our estimates are in column (1) of Table 6.

[28]For the entire sample period, MEPS also specifies military specific occupations, which are by far the least affected by the flu. We also observe that 85% of those in military occupations had been vaccinated for the flu. We drop these military specific occupation and also "unclassifiable occupations", which together consist of less than 2% of our sample.

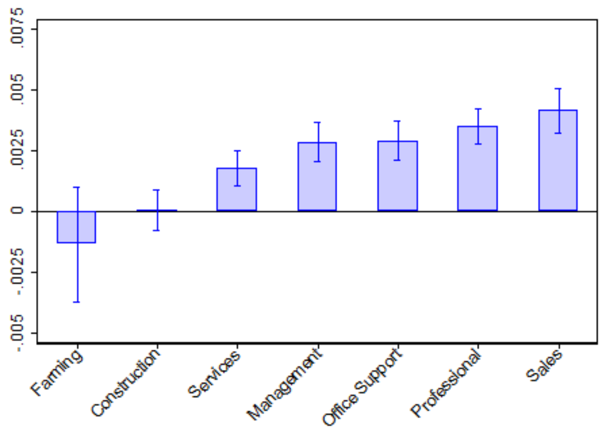
Table 6: Occupation Gradient in Flu Incidence

VARIABLES	(1) Full Sample	(2) + Controls	(3) High Incidence	(4) Low Incidence
Management	0.00287***	0.00251***	0.00383**	0.00115
	(0.000957)	(0.000963)	(0.00151)	(0.00117)
Professional	0.00349***	0.00266***	0.00418***	0.00117
	(0.000874)	(0.000893)	(0.00139)	(0.00110)
Services	0.00179**	0.00119	0.00247*	-8.48e-05
	(0.000892)	(0.000916)	(0.00143)	(0.00111)
Sales	0.00415***	0.00364***	0.00479***	0.00247*
	(0.00111)	(0.00113)	(0.00169)	(0.00145)
Office Support	0.00292***	0.00211**	0.00335**	0.000823
	(0.000985)	(0.00102)	(0.00153)	(0.00131)
Farming	-0.00134	-0.00109	0.000236	-0.00230
	(0.00292)	(0.00291)	(0.00499)	(0.00319)
Construction	6.57e-05	0.000155	0.00196	-0.00167
	(0.00102)	(0.00102)	(0.00159)	(0.00123)
Age		0.000334*	0.000820***	-0.000101
		(0.000172)	(0.000266)	(0.000218)
Age^2		-5.20e-06**	-1.11e-05***	-1.60e-08
		(2.02e-06)	(3.17e-06)	(2.52e-06)
Female		0.000496	0.00133	-0.000237
		(0.000583)	(0.000898)	(0.000728)
Never Married		-0.000489	0.000460	-0.00136
		(0.000733)	(0.00116)	(0.000884)
HH Size		-0.000183**	-0.000264*	-0.000118
		(9.26e-05)	(0.000149)	(0.000113)
N. Other Conditions		0.00176***	0.00112***	0.00239***
		(0.000242)	(0.000370)	(0.000307)
Student Status		0.00202	0.00390	0.000224
		(0.00165)	(0.00264)	(0.00197)
Constant	0.0412***	0.0359***	0.0283***	0.0203***
	(0.00256)	(0.00446)	(0.00622)	(0.00474)
Observations	2,078,384	2,078,384	1,046,949	1,031,435
R-squared	0.005	0.005	0.005	0.003
Year/Month FE	Yes	Yes	Yes	Yes
Pseudo R2	0.00470	0.00506	0.00479	0.00299

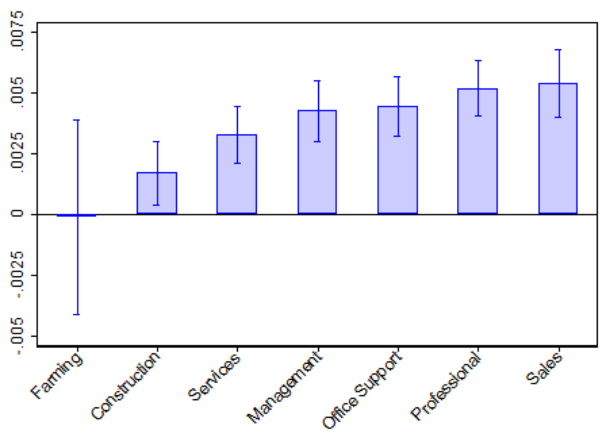
Notes: This table shows results from a linear probability model for flu incidence on MEPS data 2002-2016. Robust standard errors clustered at the individual level are in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 3: Occupation Gradient in Flu Incidence, MEPS 2002-2016

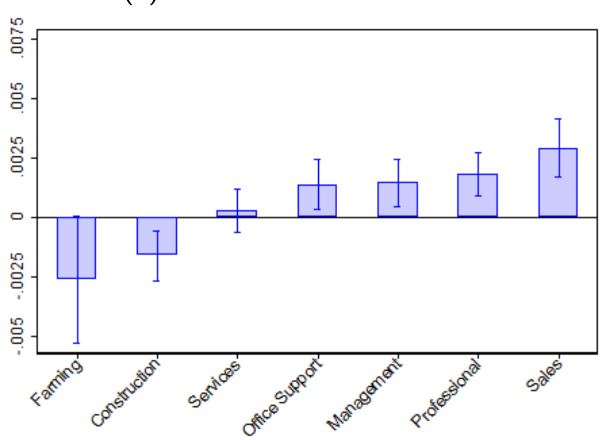
(a) Full Sample, All Years



(b) High Flu Incidence Years



(c) Low Flu Incidence Years



Notes: The occupations in MEPS are (1) management, business, and financial operations; (2) professional and related occupations; (3) service occupations; (4) sales and related occupations; (5) office and administrative support; (6) farming, fishing and forestry; (7) construction, extraction, and maintenance; and (8) production, transportation, and material moving operations. Our data is retrieved from MEPS, 2002-2016.

There is a clear gradient in flu incidence by occupations without and with individual controls, see, respectively, see column (1) and (2) of Table 6 . Focusing on the conditional specification, and given that "production, transportation, and material moving operations" has an average flu incidence of 1.49%, our estimation results implies that occupations in "sales and related occupations" are 24.4% more likely to get infected with the flu, "professional and related occupations" are 17.9% more likely to get infected with the flu, occupations in "management, business, and financial operations" are 16.8% more likely to get infected with the flu and "office and administrative support" are 14.2% more likely to get infected with the flu.[29] Further, "service occupations", "construction, extraction, and maintenance" show point estimates of flu incidence above the reference group though not significantly, respectively, 8%, 1%, while "farming, fishing and forestry" have a lower flu incidence by -7.3%. This implies an occupation-flu gradient of 34.1% between the occupation with the highest and lowest flu incidence. Our results are also shown graphically in panel (a) of Figure 3, in p.p. terms.

We reproduce the same exercise for years above and below the median annual aggregate flu incidence (1.57%). The main takeaway is that the flu incidence gradient across occupations is accentuated in years when the aggregate flu incidence is high (see column (3) in Table 6). Focusing on years above the median flu incidence year—averaging 1.16% incidence—we find that, with respect to the reference group, "production, transportation, and material moving operations", "sales and related occupations" are 41.3% more likely to be infected, "professional and related occupations" are 36.0% more likely to be infected, "management, business, and financial operations" are 33.0% more likely to be infected, "office and administrative support" are 28.9% more likely to be infected, "service occupations" are 21.3% more likely to be infected, whereas "construction, extraction, and maintenance" are 16.9% more likely to be infected, and "construction, extraction, and maintenance" are 2% more likely to be infected. The flu incidence gradient across occupations dilutes in years where the aggregate flu incidence is low (see column (4) in Table 6). Focusing on years below the median flu incidence year—averaging 1.82% incidence—we find that "sales and related occupations" are 13.6% more likely to be infected with the flu than the reference group; that is, a figure two-thirds lower than that of high incidence years. The rest of occupations are not significantly different than the reference group.

Our main results on the occupational gradient remain unaltered when we include vaccinations in our regressions (see Table A1 in Appendix A).

[29]To calculate these percentages for occupation x we use the formula $((1.49 + \beta_x)/1.49 - 1) * 100$.

3.3 Occupation-Specific Human Contact Interaction at Work

In the previous section we showed how employment status determines flu incidence and how the extent of flu incidence differs by occupations. In this section, we provide first direct evidence on how human contact interaction—a score that we construct based on occupation-specific work requirements—relates to flu incidence. Our main result is that occupation-specific human contact interaction helps determine flu incidence. We construct a measure of human contact interaction by occupation in Section 3.3.1. We assess the effects of human contact on flu incidence in Section 3.3.2.

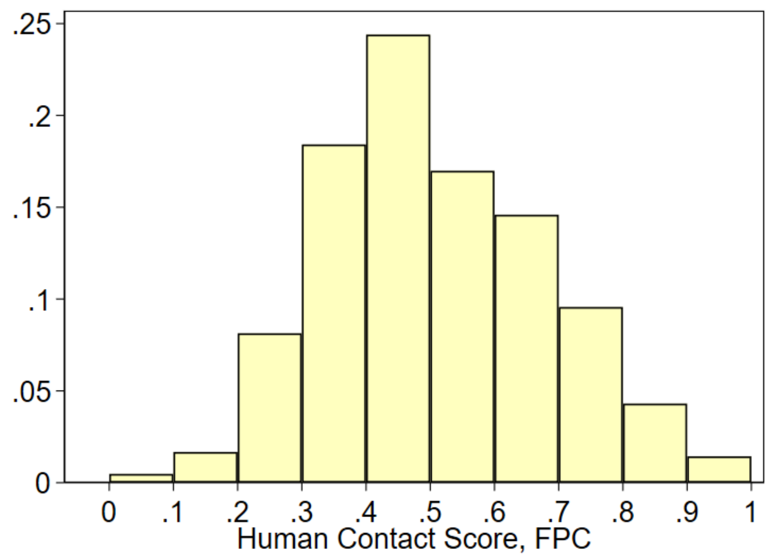
3.3.1 Measuring Human Contact by Occupation

We use the O'NET descriptors to document how occupations differ in terms of the extent of human interaction required at work. The O'NET database provides detailed information regarding work activities that require interaction with other human beings. As described in Section 2, our analysis consists of several types (descriptors) of human contact (HC) interaction. In order to take into account all the specific forms of human interaction ("Work Activities: Interacting with Others") from O'NET, we perform a principal component analysis through an eigenvalue decomposition of the covariance of all 17 descriptors in Table 2. We perform this analysis over 418 occupations.[30] That is, we summarize *all* types of occupational requirements related to human contact interaction into a human contact score, i.e., the first principal component, which we re-scale between zero and one. This first principal component, which we label as human contact score, summarizes the information on differences in human interaction by occupation with 1 denoting the maximum human interaction, and 0 denoting the minimum human interaction.

In Figure 4 we show the histogram of the re-scaled human contact score obtained across the 418 occupations in O'NET. The median occupation shows a human contact score of 0.48 and mean of the distribution is 0.51 with a standard deviation of 0.174. The lowest one percent of the distribution shows a score of 0.16 and the highest one percent of the distribution shows a score of 0.91. Notice that the distribution is asymmetric with a positively skewed pattern. On the right end of the distribution, we find that approximately 1.5% of the occupations have a score larger than a score of 0.9, and 5.1% of occupations have a score larger than a score of 0.8. On the left-end of the distribution we find that approximately less than 0.3% of the occupations show a score below 0.1, and less than 3% of the distribution of occupations shows a score below 0.2 at the bottom of the distribution where to reach 10% one has to go to a score above 0.3.

[30] The occupational codes in the O'NET are based on the SOC classification at 6 digit level. We work with the 5-digit codes by taking averages over corresponding 6-digit occupations.

Figure 4: Human Contact Score Distribution, O'NET



Notes: The human conctact score is computed by occupation as the first principal component resulting from the eigenvalue decomposition of the covariance of all 17 descriptors in Table 2. This table shows the empirical distribution of the human contact score resulting from 418 occupations (5-digit codes).

The distribution is positively skewed with a coefficient is 0.26. The kurtosis coefficient is of 2.77. Shapiro-Wilk normality tests reject normality for our human contact score.

To give a sense of which occupations require the least and the most human contact, we report in panel (a) of Table 7 the occupations with the lowest (bottom 0.25%) and highest (top 0.25%) values of our human contact score. Occupations with the least exposure to human contact interaction are mathematical science occupations, proofreaders, hunters and trappers, rock splitters, and different types of machine operators related to a variety of activities including shoe making, sewing, sawing, woodworking, metal fabricators, etc. Occupations where workers are highly exposed to the human contact interaction include chief executives, medical and health services managers, education administrators, human resources managers, training and development managers, lodging managers, and first-line supervisors of different types (sales workers, police, detectives and correctional officers, and fire fighting and prevention workers).

Further, we separately assess specific forms of human contact exposure. First, in terms of occupations, we focus on the descriptor "Performing for or Working Directly with the Public" which eludes to direct contact of job performance with others. Occupations with the least exposure to this type of human interaction are similar to those from our first principal component: "Mathematical Science Occupations" as well as all sorts of machine operators; see panel (b) of

Table 7: Occupations by Human Contact Interaction: Lowest and Highest

(a) Occupations with Highest and Lowest Human Contact Score (FPC)

Lowest	Highest
Mathematical Science Occupations, All Other (*Lowest*)	First-Line Supervisors of Sales Workers
Mathematical Technicians	Dietitians and Nutritionists
Proofreaders and Copy Markers	Urban and Regional Planners
Couriers and Messengers	Lodging Managers
Pressers, Textile, Garment, and Related Materials	Medical and Health Services Managers
Shoe Machine Operators and Tenders	Directors, Religious Activities and Education
Shoe and Leather Workers and Repairers	Education Administrators
Structural Metal Fabricators and Fitters	First-Line Supervisors of Police, Detectives and Correctional Officers
Rock Splitters, Quarry	First-Line Supervisors of Fire Fighting and Prevention Workers
Sawing Machine Setters, Operators, and Tenders	Chief Executives
Woodworking Machine Setters, Operators, and Tenders	Training and Development Managers
Hunters and Trappers	Emergency Management Directors
Sewing Machine Operators	Human Resources Managers (*Highest*)

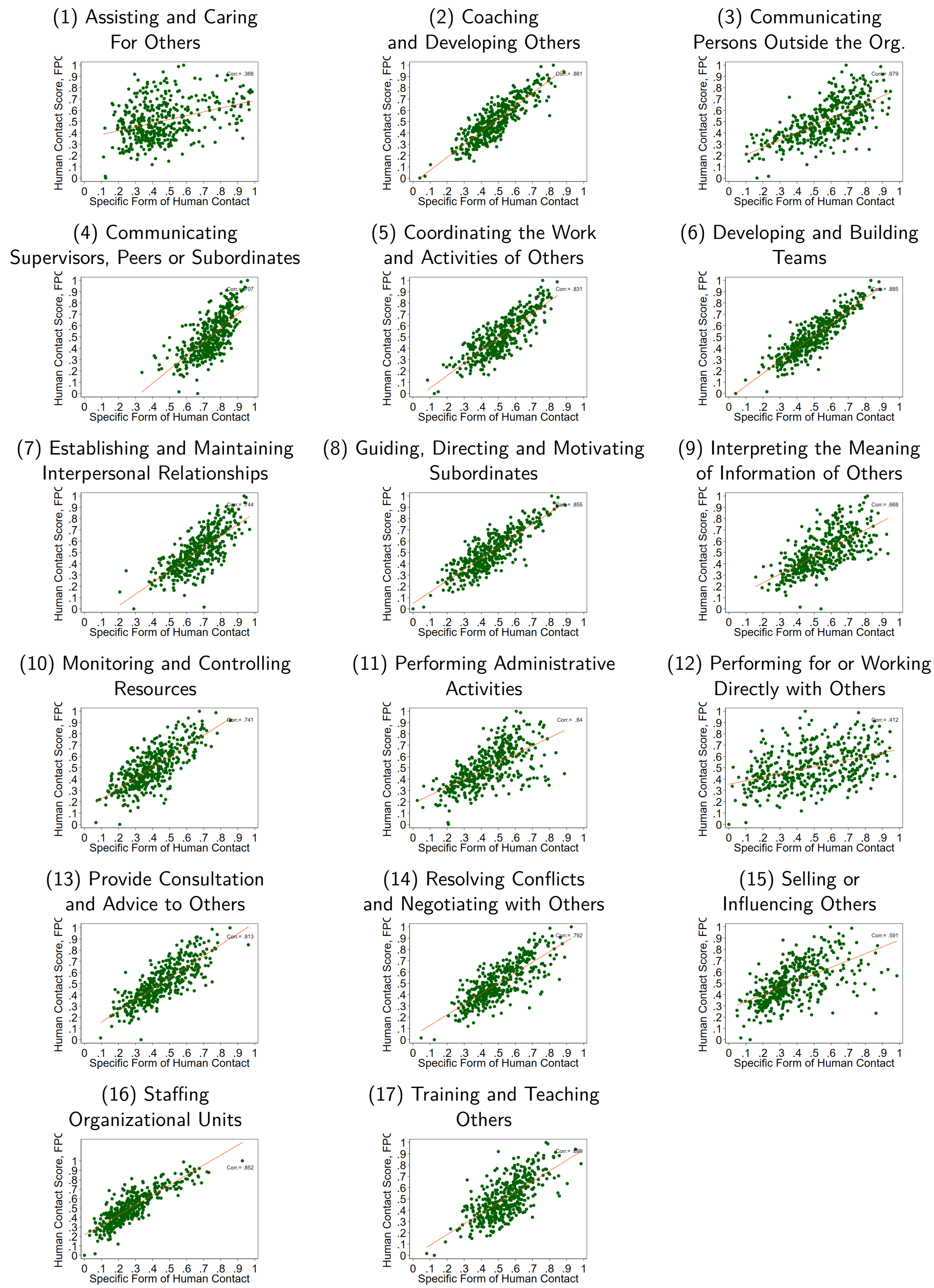
(b) Occupations with Highest and Lowest Score in "Performing for or Directly Working with the Public"

Lowest	Highest
Mathematical Technicians (Lowest)	Concierges
Mathematical Science Occupations, All Other	Police and Sheriff's Patrol Officers
Roof Bolters, Mining	Funeral Service Managers
Aerospace Engineers	Slot Supervisors
Mine Shuttle Car Operators	Gaming Supervisors
Computer Hardware Engineers	Opticians, Dispensing
Graders and Sorters, Agricultural Products	Retail Salespersons
Woodworking Machine Setters, Operators, and Tenders	Reservation and Transportation Ticket Agents and Travel Clerks
Sawing Machine Setters, Operators, and Tenders	New Accounts Clerks
Mining Machine Operators, All Other	Morticians, Undertakers, and Funeral Directors
Mine Cutting and Channeling Machine Operators	Flight Attendants (Highest)

Note: Panel (a) shows the occupations with the lowest and the highest value of our human contact score which corresponds to the first principal component of 17 descriptors in the "Work Activities: Interacting with Others" section of O'NET. Panel (b) shows the occupations with the lowest and the highest value of the normalized importance scale for one such work activities: "Performing for or Working Directly with the Public" from O'NET.

Table 7. Compared with our first principal component analysis, the landscape of occupations somewhat changes in terms of the highest exposure to this type of human interaction which now include "Flight attendants," "Morticians, Undertakers and Funeral Directors," "Clerks," "Retail salespersons," "Police and sheriff's Patrol Officers," etc. Second, we relate human contact score (FPC) to specific forms of human interaction. The human contact score is strongly and positively correlated with each specific form of human interaction, see Figure 5. The three descriptors with the highest correlations with the human contact score are "Guiding, Directing, and Motivating Subordinates" (0.854), "Coaching and Developing Others" (0.861) and "Developing and Building Teams" (0.885). The three descriptors with the lowest correlation with the first principle component are "Assisting and Caring for Others" (0.366), "Performing for or Working Directly with Public" (0.412) and "Selling or Influencing Others" (0.590). All correlations are significant.

Figure 5: Relation between Human Contact Score and Specific Forms of Human Contact



Notes: This figure shows the correlation between each of the specific forms of human contact and the human contact score (first principal component) based on the O'NET data.

3.3.2 Flu Infection by Human Contact Interaction at Work

Our O'NET human contact score can be merged to the MEPS data[31] so that we can explore the relationship between human contact interaction and flu incidence. We proceed analogously to (3), where instead of occupation fixed effects we directly use the human contact score:

$$flu_{it} = cons + \gamma \ln(HC) + \sum_{t} \omega_t \mathbf{1}_t + \sum_{x} \beta x_{it} + u_{it}, \tag{4}$$

where our parameter of interest is γ, which captures the effects of the human contact score on the probability of flu infection. Our results are in Table 8. Across all specifications, the higher is the human contact score, the higher is the probability of getting infected with the flu.

In column (1) of Table 8, we show the unconditional effects of our human contact score and the probability of getting infected with the virus. A one percent increase in the human contact score increases the probability of infection by 0.473 p.p. This result is robust to the introduction of our battery of individual controls; see column (2) of Table 8. The conditional effect of the human contact score implies an increase in the probability of infection by 0.406 p.p. We note that the effects of the individual observable characteristics are similar to those obtained when studying the effects of employment on the probability of infection; see Section 3.1. The effects of age on the probability of infection are significantly concave. Women show a higher probability of infection, though not significantly. Never married individuals show a lower probability of infection. A higher household size reduces the probability of infection. The larger is the number of other conditions also increases the probability of infection. Student status increases the probability of infection, though not significantly.

Further, we split the sample in years of high annual aggregate flu incidence (above median) and low annual aggregate flu incidence (below median); see, respectively, column (3) and (4) in Table 8. We find that the human contact score significantly increases the probability of infection by a significant 0.532 p.p. in years of high annual aggregate incidence, whereas this effect is reduced in years of low annual aggregate incidence where the human contact score increases the probability of infection by 0.282 p.p.

Finally, the effect of the human contact score is robust to firm size and industries. A potentially relevant aspect for contagion at work is the size of the firm in terms of the number of employees. Our reasoning is that larger firms can facilitate more human contact interaction compared with

[31] To merge the scores of the O'NET occupations with the MEPS occupations we use several occupation codes crosswalks and the CPS occupational employment and individual weights. More details can be found in Appendix B.

smaller firms. This would imply that the virus can spread more easily within large firms than within small firms. In this direction, the self-employed, defined in the MEPS as those that do not earn a wage but work as an entity, report respectively a smaller number of employees than the firm size reported by wage earners, 9.39 versus 145.4, and show a significantly lower probability of infection compared with the wage earners; see Section 3.1. When we control for the number of employees we find an effect of the human contact score on the probability of infection of 0.420 p.p.; see column (5) of Table 8. Note that we also find that the number of employees increases the probability of infection in a concave fashion. Further, if we control for industry fixed effects, we find that the human contact score increases the probability of infection by 0.298 p.p.; see column (6) of Table 8. We next control for individuals holding more than one job and last, restricting out attention to wage earners, we control for weekly hours worked conditional on working[32]. Our main finding remains: human contact increases the probability of infection by 0.404 p.p. and 0.479 p.p. when controlling for multiple jobs and hours worked, respectively; see columns (7) and (8) of Table 8. The coefficients on holding more jobs and hours worked are positive, yet not significant.

[32]We trim the upper and lower 1% of reported weekly hours in the MEPS, corresponding to less than 5 and more than 72 weekly hours worked.

Table 8: Occupation-Specific Human Contact at Work and Flu Incidence

VARIABLES	(1) Full Sample	(2) + Controls	(3) High Incidence	(4) Low Incidence	(5) + N. Employees	(6) + Industry	(7) More Jobs	(8) Hours Worked
$\log(HC)$	0.00473***	0.00407***	0.00532***	0.00282*	0.00420***	0.00298**	0.00404***	0.00479***
	(0.00118)	(0.00120)	(0.00186)	(0.00148)	(0.00120)	(0.00133)	(0.00120)	(0.00132)
N. Employees					0.00108*			
					(0.000642)			
N. Emp.^{2}					-0.000283**			
					(0.000126)			
Multiple Jobs							0.00106	
							(0.00100)	
Log(hours)								0.000288
								(0.000790)
Age		0.000319*	0.000803***	-0.000115	0.000335*	0.000305*	0.000317*	0.000217
		(0.000172)	(0.000265)	(0.000219)	(0.000173)	(0.000173)	(0.000172)	(0.000183)
Age^2		-5.04e-06**	-1.09e-05***	1.61e-07	-5.21e-06**	-4.89e-06**	-5.01e-06**	-3.68e-06*
		(2.02e-06)	(3.16e-06)	(2.53e-06)	(2.03e-06)	(2.03e-06)	(2.02e-06)	(2.17e-06)
Female		0.000858	0.00171**	8.87e-05	0.000856	0.000258	0.000849	0.000845
		(0.000537)	(0.000826)	(0.000673)	(0.000538)	(0.000583)	(0.000538)	(0.000590)
Never Married		-0.000472	0.000496	-0.00136	-0.000470	-0.000488	-0.000489	-0.000772
		(0.000730)	(0.00116)	(0.000880)	(0.000730)	(0.000731)	(0.000729)	(0.000769)
HH Size		-0.000187**	-0.000268*	-0.000122	-0.000191**	-0.000181*	-0.000186**	-0.000195**
		(9.25e-05)	(0.000149)	(0.000113)	(9.26e-05)	(9.27e-05)	(9.25e-05)	(9.89e-05)
N. Other Conditions		0.00178***	0.00114***	0.00240***	0.00178***	0.00175***	0.00177***	0.00169***
		(0.000243)	(0.000370)	(0.000307)	(0.000243)	(0.000243)	(0.000243)	(0.000259)
Student Status		0.00215	0.00401	0.000356	0.00213	0.00219	0.00213	0.00216
		(0.00165)	(0.00264)	(0.00197)	(0.00165)	(0.00165)	(0.00165)	(0.00170)
Constant	0.0467***	0.0405***	0.0351***	0.0229***	0.0403***	0.0372***	0.0405***	0.0431***
	(0.00260)	(0.00450)	(0.00632)	(0.00479)	(0.00450)	(0.00492)	(0.00449)	(0.00543)
Observations	2,078,384	2,078,384	1,046,949	1,031,435	2,078,384	2,078,384	2,078,384	1,838,330
R-squared	0.005	0.005	0.005	0.003	0.005	0.005	0.005	0.005
Year/Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	No	No	Yes	No	No
Pseudo R2	0.00465	0.00503	0.00476	0.00293	0.00507	0.00509	0.00504	0.00509

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table shows results from a linear probability model for flu incidence on MEPS data 2002-2016. Robust standard errors clustered at the individual level are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

4 Flu incidence by Industry

The extent of flu incidence might differ by industries in manners that escape the occupation structure. Here, we study flu incidence by industries. We follow the classification of industries in MEPS: "natural resources", "mining", "construction", "manufacturing", "wholesale and retail trade", "transportation and utilities, "information", "financial activities", "professional and business services", "education, health, and social services", "leisure and hospitality", "other services", and "public administration". We drop military and "unclassifiable" industries. Pooling all sample years, we find that the size of each of these industries in terms of employment is: "natural resources" represents 1.3% of the economy, "mining" 0.4%, "construction" 7.1%, "manufacturing" 11.3%, "wholesale and retail trade", 13.4%, "Transportation and utilities" 4.9%, "information" 2.4%, "financial activities" 6.7%, "professional and business services" 11.3%, "education, health and social services" 22.7%, "leisure and hospitality" 8.4%, "other services" 4.9% and "public administration" 5.2%.

Panel (a) of Figure 6 shows the flu incidence by industry for all our sample years. We choose "natural resources" to be our reference industry because is the industry with the lowest flu incidence, on average. Precisely, we show the β's from the following regression:

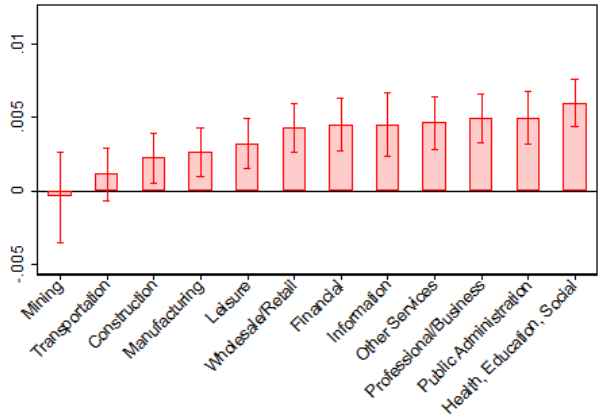
$$flu_i = cons + \sum_{ind \neq 10} \beta_{ind} \mathbf{1}_{ind} + \sum_t \omega_t \mathbf{1}_t + u_{it}. \tag{5}$$

We find that "mining" and "transportation and utilities" show no significant difference in flu incidence with respect to the reference industry. We find an industry gradient showing between 2 p.p. and 5 p.p. higher probability of being infected with the flu than the reference industry for "construction", "manufacturing", "leisure and hospitality", "wholesale and retail trade", "financial activities" and "information". The industry gradient is above 5 p.p. for "other services", "professional and business services" and "public administration". Finally, "education, health and social services" is associated with 6% higher probability of infection. As it was the case in occupations, the industry gradient is steeper in high flu incidence years (see panel (b) of Figure 6) than in low flu incidence years (see panel (c) of Figure 6).

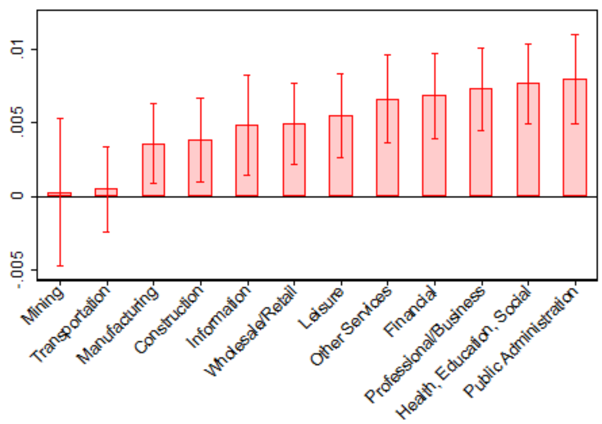
Flu and Occupations Within Industry A potential reason for the industry differentials in flu incidence is the within-industry occupation structure. In Figure 7, we show the fraction of employed individuals by occupation separately within each industry. There are substantial differences across industries in both: (1) the occupational structure and (2) in the probability of being infected with the flu by occupation.

Figure 6: Industry Gradient in Flu Incidence, MEPS 2002-2016

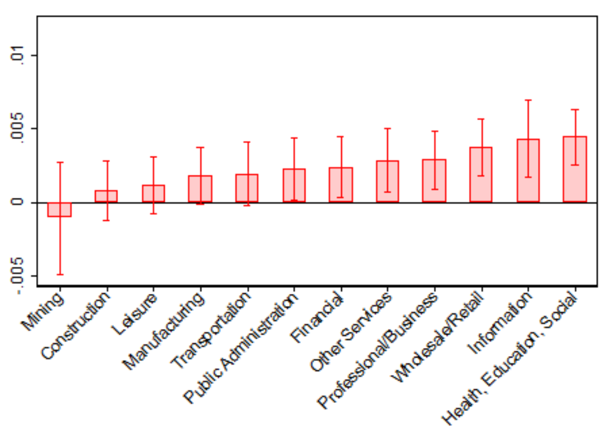
(a) Full Sample, All Years



(b) High Flu Incidence Years



(c) Low Flu Incidence Years



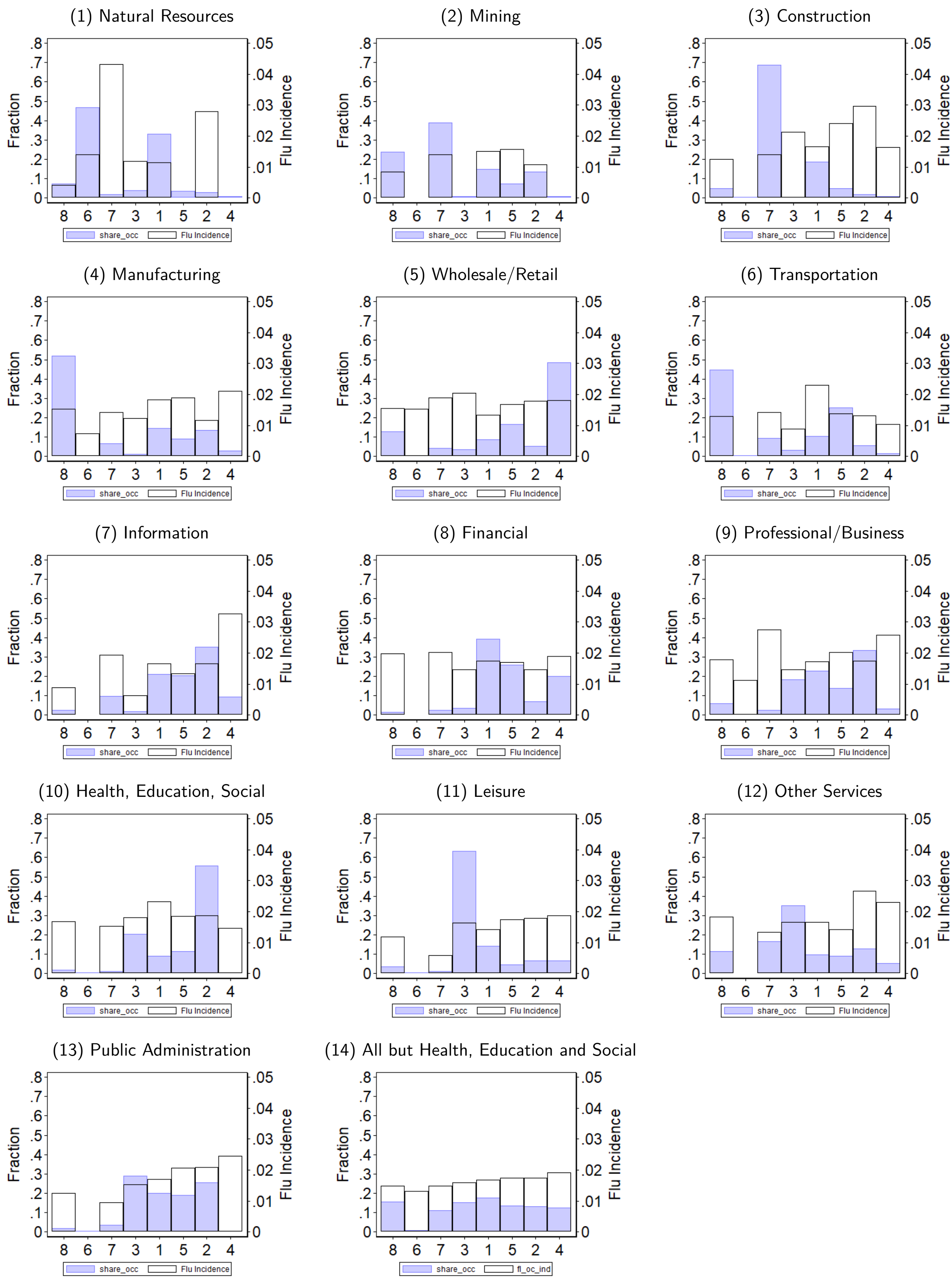
Notes: The industries in MEPS are categorized into twelve groups: (1) natural resources, (2) mining, (3) construction, (4) manufacturing, (5) wholesale and retail trade, (6) transportation and utilities, (7) information, (8) financial activities, (9) professional and business services, (10) education, health, and social services, (10) education, health, and social services, (11) leisure and hospitality, and (12) other services.

Comparing the industry with the largest flu incidence—"education, health, and social services"—and the rest of the economy we find that "education, health, and social services" have a disproportionately larger fraction of their employed in occupations that are relatively more subject to flu contagion risk. Precisely, risky occupations such as those related to management, professionals, and office support account for, respectively, 8.9%, 55.8%, and 20.5% of the total employment in "education, health, and social services," whereas these figures are 17.5%, 13.3% and 13.6% for the rest of the economy. Indeed, in the rest of the economy less risky occupations are more predominant, except for sales, the riskiest occupation, that accounts for 12.6% of total employment in the rest of the economy and for 0.3% in "education, health, and social services." The probability of infection by occupation also differs by industries. In the industry with the largest flu incidence, "education, health, and social services," we find that the probability of infection is larger in all occupations except in sales compared with the rest of the economy.[33]

In this context, we conduct a simple decomposition exercise to assess the sources of flu incidence differences across industries. We focus on a comparison between the industry with the largest flu incidence—"education, health, and social services" (EHSS)—and the rest of the economy (RoE) with, respectively, incidence levels of $flu_{EHSS} = 1.89\%$ and $flu_{RoE} = 1.66\%$. The differential incidence can be driven by the fact that "education, health, and social services" has a different occupation structure than the rest of the economy, or by the fact that the probability of being infected with the flu by occupation in "education, health, and social services" differs from the rest of the economy. To address this question, we first impose the occupation structure of "education, health, and social services" onto the other industries, keeping everything else constant. We find that occupation structure explains $\frac{(flu_{RoE}|occ.\ structure_{EHSS})-flu_{RoE}}{flu_{EHSS}-flu_{RoE}} = 20.3\%$ of the difference in flu incidence between "education, health, and social services" and the rest of the economy. Second, we impose the probably of being infected with the flu by occupation in "education, health, and social services" onto the other industries, keeping everything else constant. We find that the fact that occupations in "education, health, and social services" show a different probability of being infected with the flu explains $\frac{(flu_{RoE}|flu\ by\ occ_{EHSS})-flu_{RoE}}{flu_{EHSS}-flu_{RoE}} = 62.6\%$ of the differences in flu incidence between "education, health, and social services" and the rest of the economy. Splitting in two the joint effect of the occupation structure and the flu incidence by occupation, we find that differences in occupation structure and the differences in the probability of being infected with the flu by occupation explain, respectively, 28.9% and 71.1% of the total difference in flu incidence between "education, health, and social services" and the rest of the economy.

[33]We cannot make a comparison of flu incidence for the farming occupation since it is not present in the "education, health, and social services" industry.

Figure 7: Flu and Occupations within Industries



Notes: The horizontal axis in all panels displays the occupations with the following labels: (1) management, business, and financial operations; (2) professional and related occupations; (3) service occupations; (4) sales and related occupations; (5) office and administrative support; (6) farming, fishing and forestry; (7) construction, extraction, and maintenance; and (8) production, transportation, and material moving operations. The left vertical axis shows the fraction of individuals by occupation in a given industry. The right vertical axis shows the flu incidence by occupation in a given industry.

5 Conclusion

In this paper we exploit a rare opportunity to link occupations and flu incidence at the individual level using a nationally representative panel survey. With this data, we conduct a novel study of the effects of employment—and occupations—on actual contagion. We find sizable effects of employment on flu incidence. Further, within the employed, we find an occupation-flu gradient. In addition, we are able to positively assess a mechanism through which occupations matter for contagion: occupation-specific exposure to human contact interaction significantly determines flu incidence. We hope that our results are relevant to the understanding of the spread of the flu as well as of other infectious diseases that, as the flu, are transmitted via droplets, fomites and close human contact interaction (e.g., SARS and Covid-19).

References

Aleman, C., Busch, C., Ludwig, A., and Santaeulàlia-Llopis, R. (2020a). Evaluating the Effectiveness of Policies Against a Pandemic. Working Paper 1209, Barcelona GSE.

Aleman, C., Iorio, D., and Santaeulàlia-Llopis, R. (2020b). A Quantitative Theory of the HIV Epidemic: Education, Risky Sex and Asymmetric Learning. Working paper, Barcelona GSE.

Almagro, M. and Orane-Hutchinson, A. (2020). The determinants of the differential exposure to COVID-19 in New York city and their evolution over time. *Journal of Urban Economics*, page 103293.

Alvarez, F. E., Argente, D., and Lippi, F. (2020). A Simple Planning Problem for COVID-19 Lockdown. NBER Working Papers 26981, National Bureau of Economic Research, Inc.

Atkeson, A. (2020). What Will Be the Economic Impact of COVID-19 in the US? Rough Estimates of Disease Scenarios. Working Paper 26867, National Bureau of Economic Research.

Azzimonti, M., Fogli, A., Perri, F., and Ponder, M. (2020). Pandemic Control in ECON-EPI Networks. Working Paper 27741, National Bureau of Economic Research.

Barro, R. J., Ursúa, J. F., and Weng, J. (2020). The Coronavirus and the Great Influenza Pandemic: Lessons from the "Spanish Flu" for the Coronavirus's Potential Effects on Mortality and Economic Activity. Working Paper 26866, National Bureau of Economic Research.

Belongia, E. A., Simpson, M. D., King, J. P., Sundaram, M. E., Kelley, N. S., Osterholm, M. T., and McLean, H. Q. (2016). Variable Influenza Vaccine Effectiveness by Subtype: A Systematic Review and Meta-analysis of Test-Negative Design Studies. *Lancet*, 8:942–951.

Berger, D., Herkenhoff, K., and Mongey, S. (2020). An SEIR Infectious Disease Model with Testing and Conditional Quarantine. Working paper, BFI.

Casares, M. and Khan, H. (2020). A Dynamic Model of COVID-19: Contagion and Implications of Isolation Enforcement. Technical report.

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020). The Macroeconomics of Epidemics. NBER Working Papers 26882, National Bureau of Economic Research, Inc.

Fajgelbaum, P. D., Khandelwal, A., Kim, W., Mantovani, C., and Schaal, E. (2020). Optimal Lockdown in a Commuting Network. Working paper, Barcelona GSE.

Farboodi, M., Jarosch, G., and Shimer, R. (2020). Internal and External Effects of Social Distancing in a Pandemic. Working Paper 27059, National Bureau of Economic Research.

Ferguson, N., C.-D. C. S. e. a. (2005). Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(209–214).

Garibaldi, P., Moen, E. R., and Pissarides, C. A. (2020). Modelling Contacts and Transitions in the SIR Epidemics Model. *Covid Economics: Vetted and Real-Time Papers*, 5:1–20.

Glover, A., Heathcote, J., Krueger, D., and Ríos-Rull, J.-V. (2020). Health versus Wealth: On the Distributional Effects of Controlling a Pandemic. Working Paper 27046, National Bureau of Economic Research.

Halloran, M. E., Ferguson, N. M., Eubank, S., Longini, I. M., Cummings, D. A., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T. C., Wagener, D., Beckman, R., Kadau, K., Barrett, C., Macken, C. A., Burke, D. S., and Cooley, P. (2008). Modeling Targeted Layered Containment of an Influenza Pandemic in the United States. *Proc Natl Acad Sci U S A*, 105(12):4639–4644.

Iorio, D. and Santaeulàlia-Llopis, R. (2016). Education, HIV Status, and Risky Sexual Behavior: How Much Does the Stage of the HIV Epidemic Matter? Working paper, Barcelona GSE.

Kaplan, G., Moll, B., and Violante, G. (2020). The Great Lockdown and the Big Stimulus: Tracing the Pandemic Possibility Frontier for the U.S. Working paper, National Bureau of Economic Research.

Klepac, P., Kucharski, A. J., Conlan, A. J., Kissler, S., Tang, M. L., Fry, H., and Gog, J. R. (2020). Contacts in Context: Large-Scale Setting-Specific Social Mixing Matrices from the BBC Pandemic Project. Technical report.

Mongey, S., Pilossoph, L., and Weinberg, A. (2020). Which Workers Bear the Burden of Social Distancing Policies? Working Paper 27085, National Bureau of Economic Research.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., and Edmunds, W. J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine*, 5(3):1–1.

Obiols-Homs, F. (2020). Precaution, Social Distancing and Tests in a Model of Epidemic Disease. Working paper, Barcelona GSE.

Piguillem, F. and Shi, L. (2020). Optimal COVID-19 Quarantine and Testing Policies. Technical report.

Appendix

A Occupational Gradient Including Vaccination

Table A1: Occupation Gradient in Flu Incidence: Including Vaccinations

VARIABLES	(1) Full Sample	(2) + Controls	(3) High Incidence	(4) Low Incidence
Flu Shot	0.00108**	0.000830	0.00153*	0.000204
	(0.000544)	(0.000551)	(0.000852)	(0.000678)
Management	0.00281***	0.00248**	0.00344**	0.00147
	(0.000973)	(0.000979)	(0.00154)	(0.00118)
Professional	0.00334***	0.00257***	0.00383***	0.00133
	(0.000893)	(0.000911)	(0.00142)	(0.00111)
Services	0.00180**	0.00122	0.00225	0.000180
	(0.000905)	(0.000929)	(0.00146)	(0.00112)
Sales	0.00424***	0.00375***	0.00460***	0.00287*
	(0.00113)	(0.00114)	(0.00171)	(0.00147)
Office Support	0.00254**	0.00177*	0.00284*	0.000655
	(0.000993)	(0.00103)	(0.00155)	(0.00130)
Farming	-0.000918	-0.000686	0.000530	-0.00184
	(0.00300)	(0.00299)	(0.00512)	(0.00329)
Construction	0.000151	0.000214	0.00166	-0.00127
	(0.00103)	(0.00103)	(0.00161)	(0.00123)
Age		0.000359**	0.000833***	-6.60e-05
		(0.000174)	(0.000270)	(0.000218)
Age^2		-5.55e-06***	-1.14e-05***	-4.51e-07
		(2.04e-06)	(3.21e-06)	(2.51e-06)
Female		0.000497	0.00117	-9.59e-05
		(0.000593)	(0.000911)	(0.000741)
Never Married		-0.000453	0.000492	-0.00134
		(0.000747)	(0.00118)	(0.000904)
HH Size		-0.000184*	-0.000251*	-0.000129
		(9.39e-05)	(0.000152)	(0.000114)
N. Other Conditions		0.00171***	0.00111***	0.00231***
		(0.000247)	(0.000376)	(0.000312)
Student Status		0.00188	0.00355	0.000256
		(0.00167)	(0.00267)	(0.00200)
Constant	0.0410***	0.0353***	0.0280***	0.0195***
	(0.00259)	(0.00449)	(0.00629)	(0.00472)
Observations	2,029,552	2,029,552	1,025,716	1,003,836
R-squared	0.005	0.005	0.005	0.003
Year/Month FE	Yes	Yes	Yes	Yes
Pseudo R2	0.00472	0.00508	0.00483	0.00297

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

B Merging Human Contact Scores of O'NET to MEPS occupations

The O'NET Work Activities files contains 968 different occupational codes and 774 5-digit SOC codes. Some occupational codes in the O'NET compared to the SOC classification are more detailed. e.g. the code 25-1000 for post-secondary teachers contains 36 different occupational codes in the O'NET. For these, we take simple averages of the importance of each of the work activities in the file. We are finally left with 421 occupations. Using a crosswalk, we merge these occupations to the CPS data, which contains census occupational codes. Our CPS sample spans the years 1976 to 2019 and includes 3,728,362 observations.

We then use the crosswalk provided by the MEPS[34] to convert detailed census codes into the wide occupational groups in the MEPS. We show in Table B1 the employment shares of each MEPS occupation in the CPS, together with the average and the standard deviation of our measure of the Human Contact Score (based on the First principle component from O'NET). We then merge the average Human contact score to our MEPS data.

Table B1: Human Contact Scores of MEPS Occupations

MEPS Occupation	Employment share	Human Score Average (CPS)	Human Score Std (CPS)
1	0.15	.732925	.0987343
2	0.19	.5797626	.1328877
3	0.16	.4896034	.1387376
4	0.11	.5925368	.1693563
5	0.13	.4402962	.1132504
6	0.0	.3861701	.0627149
7	0.09	.4347584	.1159099
8	0.1	.3587206	.1170752

Notes: This table shows the employment shares of each MEPS occupation in the CPS sample, together with a Human Score average and standard deviation. The Human Score index is the first principle component of detailed activities from the O'NET data.

[34] The crosswalk can be found at https://meps.ahrq.gov/data_stats/download_data/pufs/ind_occ/occ3.shtml